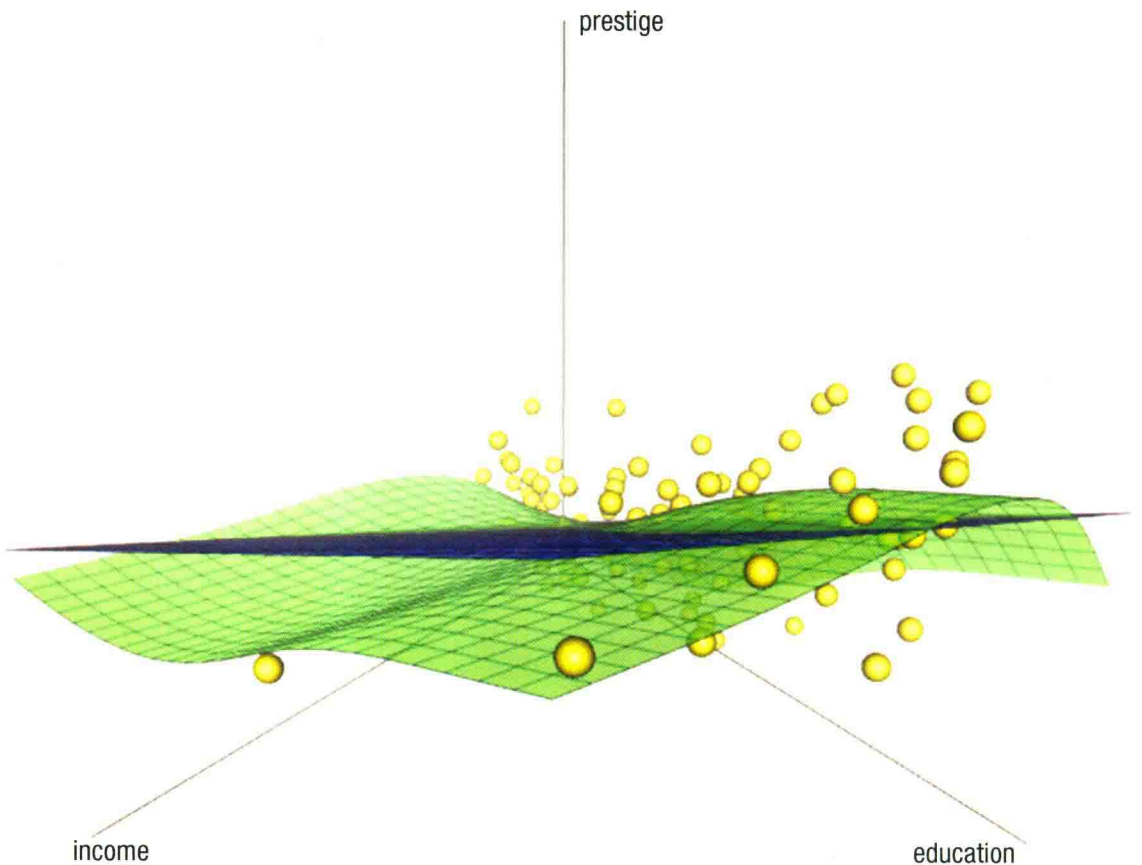


3^{EDITION}

APPLIED REGRESSION ANALYSIS & GENERALIZED LINEAR MODELS



John Fox

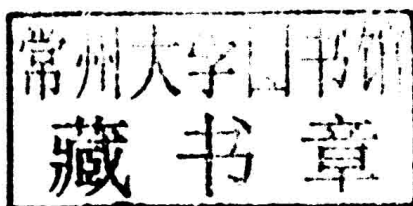


THIRD EDITION

APPLIED REGRESSION ANALYSIS and GENERALIZED LINEAR MODELS

John Fox

McMaster University



 **SAGE**

Los Angeles | London | New Delhi
Singapore | Washington DC | Boston



Los Angeles | London | New Delhi
Singapore | Washington DC | Boston

FOR INFORMATION:

SAGE Publications, Inc.
2455 Teller Road
Thousand Oaks, California 91320
E-mail: order@sagepub.com

SAGE Publications Ltd.
1 Oliver's Yard
55 City Road
London EC1Y 1SP
United Kingdom

SAGE Publications India Pvt. Ltd.
B 1/I 1 Mohan Cooperative Industrial Area
Mathura Road, New Delhi 110 044
India

SAGE Publications Asia-Pacific Pte. Ltd.
3 Church Street
#10-04 Samsung Hub
Singapore 049483

Acquisitions Editor: Vicki Knight
Associate Digital Content Editor: Katie Bierach
Editorial Assistant: Yvonne McDuffee
Production Editor: Kelly DeRosa
Copy Editor: Gillian Dickens
Typesetter: C&M Digitals (P) Ltd.
Proofreader: Jennifer Grubba
Cover Designer: Anupama Krishnan
Marketing Manager: Nicole Elliott

Copyright © 2016 by SAGE Publications, Inc.

All rights reserved. No part of this book may be reproduced or utilized in any form or by any means, electronic or mechanical, including photocopying, recording, or by any information storage and retrieval system, without permission in writing from the publisher.

Cataloging-in-Publication Data is available for this title from the Library of Congress.

ISBN 978-1-4522-0566-3

Printed in the United States of America

15 16 17 18 19 10 9 8 7 6 5 4 3 2 1

THIRD EDITION

APPLIED REGRESSION ANALYSIS and **GENERALIZED LINEAR MODELS**

For Bonnie and Jesse (yet again)

Brief Contents

Preface	xv
About the Author	xxiv
1. Statistical Models and Social Science	1
I. DATA CRAFT	12
2. What Is Regression Analysis?	13
3. Examining Data	28
4. Transforming Data	55
II. LINEAR MODELS AND LEAST SQUARES	81
5. Linear Least-Squares Regression	82
6. Statistical Inference for Regression	106
7. Dummy-Variable Regression	128
8. Analysis of Variance	153
9. Statistical Theory for Linear Models*	202
10. The Vector Geometry of Linear Models*	245
III. LINEAR-MODEL DIAGNOSTICS	265
11. Unusual and Influential Data	266
12. Diagnosing Non-Normality, Nonconstant Error Variance, and Nonlinearity	296
13. Collinearity and Its Purported Remedies	341
IV. GENERALIZED LINEAR MODELS	369
14. Logit and Probit Models for Categorical Response Variables	370
15. Generalized Linear Models	418

V. EXTENDING LINEAR AND GENERALIZED LINEAR MODELS	473
16. Time-Series Regression and Generalized Least Squares*	474
17. Nonlinear Regression	502
18. Nonparametric Regression	528
19. Robust Regression*	586
20. Missing Data in Regression Models	605
21. Bootstrapping Regression Models	647
22. Model Selection, Averaging, and Validation	669
VI. MIXED-EFFECTS MODELS	699
23. Linear Mixed-Effects Models for Hierarchical and Longitudinal Data	700
24. Generalized Linear and Nonlinear Mixed-Effects Models	743
Appendix A	759
References	762
Author Index	773
Subject Index	777
Data Set Index	791

Preface

Linear models, their variants, and extensions—the most important of which are *generalized* linear models—are among the most useful and widely used statistical tools for social research. This book aims to provide an accessible, in-depth, modern treatment of regression analysis, linear models, generalized linear models, and closely related methods.

The book should be of interest to students and researchers in the social sciences. Although the specific choice of methods and examples reflects this readership, I expect that the book will prove useful in other disciplines that employ regression models for data analysis and in courses on applied regression and generalized linear models where the subject matter of applications is not of special concern.

I have endeavored to make the text as accessible as possible (but no more accessible than possible—i.e., I have resisted watering down the material unduly). With the exception of four chapters, several sections, and a few shorter passages, the prerequisite for reading the book is a course in basic applied statistics that covers the elements of statistical data analysis and inference. To the extent that I could without doing violence to the material, I have tried to present even relatively advanced topics (such as methods for handling missing data and bootstrapping) in a manner consistent with this prerequisite.

Many topics (e.g., logistic regression in Chapter 14) are introduced with an example that motivates the statistics or (as in the case of bootstrapping, in Chapter 21) by appealing to familiar material. The general mode of presentation is from the specific to the general: Consequently, simple and multiple linear regression are introduced before the general linear model, and linear, logit, and probit models are introduced before generalized linear models, which subsume all the previous topics. Indeed, I could start with the generalized linear mixed-effects model (GLMM), described in the final chapter of the book, and develop all these other topics as special cases of the GLMM, but that would produce a much more abstract and difficult treatment (cf., e.g., Stroup, 2013).

The exposition of regression analysis starts (in Chapter 2) with an elementary discussion of nonparametric regression, developing the notion of regression as a conditional average—in the absence of restrictive assumptions about the nature of the relationship between the response and explanatory variables. This approach begins closer to the data than the traditional starting point of linear least-squares regression and should make readers skeptical about glib assumptions of linearity, constant variance, and so on.

More difficult chapters and sections are marked with asterisks. These parts of the text can be omitted without loss of continuity, but they provide greater understanding and depth, along with coverage of some topics that depend on more extensive mathematical or statistical background. I do not, however, wish to exaggerate the background that is required for this “more difficult” material: All that is necessary is some exposure to matrices, elementary linear algebra, elementary differential calculus, and some basic ideas from probability and mathematical statistics. Appendices to the text provide the background required for understanding the more advanced material.

All chapters include summary information in boxes interspersed with the text and at the end of the chapter, and most conclude with recommendations for additional reading. You will find theoretically focused exercises at the end of most chapters, some extending the material in the text. More difficult, and occasionally challenging, exercises are marked with asterisks. In addition, data-analytic exercises for each chapter are available on the website for the book, along with the associated data sets.

What Is New in the Third Edition?

The first edition of this book, published by Sage in 1997 and entitled *Applied Regression, Linear Models, and Related Methods*, originated in my 1984 text *Linear Statistical Models and Related Methods* and my 1991 monograph *Regression Diagnostics*. The title of the 1997 edition reflected a change in organization and emphasis: I thoroughly reworked the book, removing some topics and adding a variety of new material. But even more fundamentally, the book was extensively rewritten. It was a new and different book from my 1984 text.

The second edition had a (slightly) revised title, making reference to “generalized linear models” rather than to “linear models” (and dropping the reference to “related methods” as unnecessary), reflecting another change in emphasis. I retain that title for the third edition. There was quite a bit of new material in the second edition, and some of the existing material was reworked and rewritten, but the general level and approach of the book was similar to the first edition, and most of the material in the first edition, especially in Parts I through III (see below), was preserved in the second edition. I was gratified by the reception of the first and second editions of this book by reviewers and other readers; although I felt the need to bring the book up to date and to improve it in some respects, I also didn’t want to “fix what ain’t broke.”

The second edition introduced a new chapter on generalized linear models, greatly augmenting a very brief section on this topic in the first edition. What were previously sections on time-series regression, nonlinear regression, nonparametric regression, robust regression, and bootstrapping became separate chapters, many with extended treatments of their topics. I added a chapter on missing data and another on model selection, averaging, and validation (incorporating and expanding material on model validation from the previous edition).

Although I have made small changes throughout the text, the principal innovation in the third edition is the introduction of a new section on mixed-effects models for hierarchical and longitudinal data, with chapters on linear mixed-effects models and on nonlinear and generalized linear mixed-effects models (Chapters 23 and 24). These models are used increasingly in social research and I thought that it was important to incorporate them in this text. There is also a revised presentation of analysis of variance models in Chapter 8, which includes a simplified treatment, allowing readers to skip the more complex aspects of the topic, if they wish; an introduction to instrumental-variables estimation and two-stage least squares in Chapter 9; and a brief consideration of design-based inference for statistical models fit to data from complex survey samples in Chapter 15.

As in the second edition, the appendices to the book (with the exception of Appendix A on notation) are on the website for the book. In addition, data-analytic exercises and data sets from the book are on the website.

Synopsis

Chapter 1 discusses the role of statistical data analysis in social science, expressing the point of view that statistical models are essentially descriptive, not direct (if abstract)

representations of social processes. This perspective provides the foundation for the data-analytic focus of the text.

Part I: Data Craft

The first part of the book consists of preliminary material:¹

Chapter 2 introduces the notion of regression analysis as tracing the conditional distribution of a response variable as a function of one or several explanatory variables. This idea is initially explored “nonparametrically,” in the absence of a restrictive statistical model for the data (a topic developed more extensively in Chapter 18).

Chapter 3 describes a variety of graphical tools for examining data. These methods are useful both as a preliminary to statistical modeling and to assist in the diagnostic checking of a model that has been fit to data (as discussed, e.g., in Part III).

Chapter 4 discusses variable transformation as a solution to several sorts of problems commonly encountered in data analysis, including skewness, nonlinearity, and nonconstant spread.

Part II: Linear Models and Least Squares

The second part, on linear models fit by the method of least squares, along with Part III on diagnostics and Part IV on generalized linear models, comprises the heart of the book:

Chapter 5 discusses linear least-squares regression. Linear regression is the prototypical linear model, and its direct extension is the subject of Chapters 7 to 10.

Chapter 6, on statistical inference in regression, develops tools for testing hypotheses and constructing confidence intervals that apply generally to linear models. This chapter also introduces the basic methodological distinction between empirical and structural relationships—a distinction central to understanding causal inference in nonexperimental research.

Chapter 7 shows how “dummy variables” can be employed to extend the regression model to qualitative explanatory variables (or “factors”). Interactions among explanatory variables are introduced in this context.

Chapter 8, on analysis of variance models, deals with linear models in which all the explanatory variables are factors.

*Chapter 9** develops the statistical theory of linear models, providing the foundation for much of the material in Chapters 5 to 8 along with some additional, and more general, results. This chapter also includes an introduction to instrumental-variables estimation and two-stage least squares.

*Chapter 10** applies vector geometry to linear models, allowing us literally to visualize the structure and properties of these models. Many topics are revisited from the geometric perspective, and central concepts—such as “degrees of freedom”—are given a natural and compelling interpretation.

¹I believe that it was Michael Friendly of York University who introduced me to the term *data craft*, a term that aptly characterizes the content of this section and, indeed, of the book more generally.

Part III: Linear-Model Diagnostics

The third part of the book describes “diagnostic” methods for discovering whether a linear model fit to data adequately represents the data. Methods are also presented for correcting problems that are revealed:

Chapter 11 deals with the detection of unusual and influential data in linear models.

Chapter 12 describes methods for diagnosing a variety of problems, including non-normally distributed errors, nonconstant error variance, and nonlinearity. Some more advanced material in this chapter shows how the method of maximum likelihood can be employed for selecting transformations.

Chapter 13 takes up the problem of collinearity—the difficulties for estimation that ensue when the explanatory variables in a linear model are highly correlated.

Part IV: Generalized Linear Models

The fourth part of the book is devoted to generalized linear models, a grand synthesis that incorporates the linear models described earlier in the text along with many of their most important extensions:

Chapter 14 takes up linear-like logit and probit models for qualitative and ordinal categorical response variables. This is an important topic because of the ubiquity of categorical data in the social sciences (and elsewhere).

Chapter 15 describes the generalized linear model, showing how it encompasses linear, logit, and probit models along with statistical models (such as Poisson and gamma regression models) not previously encountered in the text. The chapter includes a treatment of diagnostic methods for generalized linear models, extending much of the material in Part III, and ends with an introduction to inference for linear and generalized linear models in complex survey samples.

Part V: Extending Linear and Generalized Linear Models

The fifth part of the book discusses important extensions of linear and generalized linear models. In selecting topics, I was guided by the proximity of the methods to linear and generalized linear models and by the promise that these methods hold for data analysis in the social sciences. The methods described in this part of the text are given introductory—rather than extensive—treatments. My aim in introducing these relatively advanced topics is to provide (1) enough information so that readers can begin to use these methods in their research and (2) sufficient background to support further work in these areas should readers choose to pursue them. To the extent possible, I have tried to limit the level of difficulty of the exposition, and only Chapter 19 on robust regression is starred in its entirety (because of its essential reliance on basic calculus).

Chapter 16 describes time-series regression, where the observations are ordered in time and hence cannot usually be treated as statistically independent. The chapter introduces the method of generalized least squares, which can take account of serially correlated errors in regression.

Chapter 17 takes up nonlinear regression models, showing how some nonlinear models can be fit by linear least squares after transforming the model to linearity, while other, fundamentally nonlinear, models require the method of nonlinear least squares. The chapter includes treatments of polynomial regression and regression splines, the latter closely related to the topic of the subsequent chapter.

Chapter 18 introduces nonparametric regression analysis, which traces the dependence of the response on the explanatory variables in a regression without assuming a particular functional form for their relationship. This chapter contains a discussion of generalized nonparametric regression, including generalized additive models.

Chapter 19 describes methods of robust regression analysis, which are capable of automatically discounting unusual data.

Chapter 20 discusses missing data, explaining the potential pitfalls lurking in common approaches to missing data, such as complete-case analysis, and describing more sophisticated methods, such as multiple imputation of missing values. This is an important topic because social science data sets are often characterized by a large proportion of missing data.

Chapter 21 introduces the “bootstrap,” a computationally intensive simulation method for constructing confidence intervals and hypothesis tests. In its most common nonparametric form, the bootstrap does not make strong distributional assumptions about the data, and it can be made to reflect the manner in which the data were collected (e.g., in complex survey sampling designs).

Chapter 22 describes methods for model selection, model averaging in the face of model uncertainty, and model validation. Automatic methods of model selection and model averaging, I argue, are most useful when a statistical model is to be employed for prediction, less so when the emphasis is on interpretation. Validation is a simple method for drawing honest statistical inferences when—as is commonly the case—the data are employed both to select a statistical model and to estimate its parameters.

Part VI: Mixed-Effects Models

Part VI, new to the third edition of the book, develops linear, generalized linear, and nonlinear mixed-effects models for clustered data, extending regression models for independent observations covered earlier in the text. As in Part V, my aim is to introduce readers to the topic, providing a basis for applying these models in practice as well as for reading more extensive treatments of the subject. As mentioned earlier in this preface, mixed-effects models are in wide use in the social sciences, where they are principally applied to hierarchical and longitudinal data.

Chapter 23 introduces linear mixed-effects models and describes the fundamental issues that arise in the analysis of clustered data through models that incorporate random effects. Illustrative applications include both hierarchical and longitudinal data.

Chapter 24 describes generalized linear mixed-effects models for non-normally distributed response variables, such as logistic regression for a dichotomous response, and Poisson and related regression models for count data. The chapter also introduces nonlinear mixed-effects models for fitting fundamentally nonlinear equations to clustered data.

Appendices

Several appendices provide background, principally—but not exclusively—for the starred portions of the text. With the exception of Appendix A, which is printed at the back of the book, all the appendices are on the website for the book.

Appendix A describes the notational conventions employed in the text.

Appendix B provides a basic introduction to matrices, linear algebra, and vector geometry, developing these topics from first principles. Matrices are used extensively in statistics, including in the starred portions of this book. Vector geometry provides the basis for the material in Chapter 10 on the geometry of linear models.

Appendix C reviews powers and logs and the geometry of lines and planes, introduces elementary differential and integral calculus, and shows how, employing matrices, differential calculus can be extended to several independent variables. Calculus is required for some starred portions of the text—for example, the derivation of least-squares and maximum-likelihood estimators. More generally in statistics, calculus figures prominently in probability theory and in optimization problems.

Appendix D provides an introduction to the elements of probability theory and to basic concepts of statistical estimation and inference, including the essential ideas of Bayesian statistical inference. The background developed in this appendix is required for some of the material on statistical inference in the text and for certain other topics, such as multiple imputation of missing data and model averaging.

Computing

Nearly all the examples in this text employ real data from the social sciences, many of them previously analyzed and published. The online exercises that involve data analysis also almost all use real data drawn from various areas of application. I encourage readers to analyze their own data as well.

The data sets for examples and exercises can be downloaded free of charge via the World Wide Web; point your web browser at www.sagepub.com/fox3e. Appendices and exercises are distributed as portable document format (PDF) files.

I occasionally comment in passing on computational matters, but the book generally ignores the finer points of statistical computing in favor of methods that are computationally simple. I feel that this approach facilitates learning. Thus, for example, linear least-squares coefficients are obtained by solving the normal equations formed from sums of squares and products of the variables rather than by a more numerically stable method. Once basic techniques are absorbed, the data analyst has recourse to carefully designed programs for statistical computations.

I think that it is a mistake to tie a general discussion of linear and related statistical models too closely to particular software. Any reasonably capable statistical software will do almost everything described in this book. My current personal choice of statistical software, both for research and for teaching, is R—a free, open-source implementation of the S statistical programming language and computing environment (Ihaka & Gentleman, 1996; R Core Team, 2014). R is now the dominant statistical software among statisticians; it is used increasingly in the social sciences but is by no means dominant there. I have coauthored a separate book (Fox & Weisberg, 2011) that provides a general introduction to R and that describes its use in applied regression analysis.

Reporting Numbers in Examples

A note on how numbers are reported in the data analysis examples: I typically show numbers to four or five significant digits in tables and in the text. This is greater precision than is usually desirable in research reports, where showing two or three significant digits makes the results more digestible by readers. But my goal in this book is generally to allow the reader to reproduce the results shown in examples. In many instances, numbers in examples are computed from each other, rather than being taken directly from computer output; in these instances, a reader comparing the results in the text to those in computer output may encounter small differences, usually of one unit in the last decimal place.

To Readers, Students, and Instructors

I have used the material in this book and its predecessors for two types of courses (along with a variety of short courses and lectures):

- I cover the unstarred sections of Chapters 1 to 8, 11 to 15, 20, and 22 in a one-semester course for social science graduate students (at McMaster University in Hamilton, Ontario, Canada) who have had (at least) a one-semester introduction to statistics at the level of Moore, Notz, and Fligner (2013). The outline of this course is as follows:

Week	Topic	Reading (Chapter)
1	Introduction to the course and to regression	1, 2
2	Examining and transforming data	3, 4
3	Linear least-squares regression	5
4	Statistical inference for regression	6
5	Dummy-variable regression and analysis of variance	7, 8
6	Review: Through dummy regression	
7	Diagnostics I: Unusual and influential data	11
8	Diagnostics II: Nonlinearity and other ills	12
9	Diagnostics III: Collinearity and model selection	13, 22
10	Logit and probit models for dichotomous data	14
11	Logit and probit models for polytomous data	14
12	Generalized linear models	15
13	Missing data	20
14	Review: From analysis of variance	

These readings are supplemented by selections from *An R Companion to Applied Regression, Second Edition* (Fox & Weisberg, 2011). Students complete required weekly homework assignments, which focus primarily on data analysis. Homework is collected and corrected but not

graded. I distribute answers after the homework is collected. There are midterm and final take-home exams (after the review classes), also focused on data analysis.

- I used the material in the predecessors of Chapters 1 to 15 and the several appendices for a two-semester course for social science graduate students (at York University in Toronto) with similar statistical preparation. For this second, more intensive, course, background topics (such as linear algebra) were introduced as required and constituted about one fifth of the course. The organization of the course was similar to the first one.

Both courses include some treatment of statistical computing, with more information on programming in the second course. For students with the requisite mathematical and statistical background, it should be possible to cover almost all the text in a reasonably paced two-semester course.

In learning statistics, it is important for the reader to participate actively, both by working through the arguments presented in the book and—even more important—by applying methods to data. Statistical data analysis is a *craft*, and, like any craft, developing proficiency requires effort and practice. Reworking examples is a good place to start, and I have presented illustrations in such a manner as to facilitate reanalysis and further analysis of the data.

Where possible, I have relegated formal “proofs” and derivations to exercises, which nevertheless typically provide some guidance to the reader. I believe that this type of material is best learned constructively. As well, including too much algebraic detail in the body of the text invites readers to lose the statistical forest for the mathematical trees. You can decide for yourself (or your students) whether or not to work the theoretical exercises. It is my experience that some people feel that the process of working through derivations cements their understanding of the statistical material, while others find this activity tedious and pointless. Some of the theoretical exercises, marked with asterisks, are comparatively difficult. (Difficulty is assessed relative to the material in the text, so the threshold is higher in starred sections and chapters.)

In preparing the data-analytic exercises, I have tried to find data sets of some intrinsic interest that embody a variety of characteristics. In many instances, I try to supply some direction in the data-analytic exercises, but—like all real-data analysis—these exercises are fundamentally open-ended. It is therefore important for instructors to set aside time to discuss data-analytic exercises in class, both before and after students tackle them. Although students often miss important features of the data in their initial analyses, this experience—properly approached and integrated—is an unavoidable part of learning the craft of data analysis.

A few exercises, marked with pound-signs (#) are meant for “hand” computation. Hand computation (i.e., with a calculator) is tedious, and is practical only for unrealistically small problems, but it sometimes serves to make statistical procedures more concrete (and increases our admiration for our pre-computer-era predecessors). Similarly, despite the emphasis in the text on analyzing real data, a small number of exercises generate simulated data to clarify certain properties of statistical methods.

I struggled with the placement of cross-references to exercises and to other parts of the text, trying brackets [too distracting!], marginal boxes (too imprecise), and finally settling on traditional footnotes.² I suggest that you ignore both the cross-references and the other footnotes on first reading of the text.³

Finally, a word about style: I try to use the first person singular—“I”—when I express opinions. “We” is reserved for you—the reader—and I.

²Footnotes are a bit awkward, but you don’t have to read them.

³Footnotes other than cross-references generally develop small points and elaborations.

Acknowledgments

Many individuals have helped me in the preparation of this book.

I am grateful to the York University Statistical Consulting Service study group, which read, commented on, and corrected errors in the manuscript, both of the previous edition of the book and of the new section on mixed-effects models introduced in this edition.

A number of friends and colleagues donated their data for illustrations and exercises—implicitly subjecting their research to scrutiny and criticism.

Several individuals contributed to this book by making helpful comments on it and its predecessors (Fox, 1984, 1997, 2008): Patricia Ahmed, University of Kentucky; Robert Andersen; A. Alexander Beaujean, Baylor University; Ken Bollen; John Brehm, University of Chicago; Gene Denzel; Shirley Dowdy; Michael Friendly; E. C. Hedberg, NORC at the University of Chicago; Paul Herzberg; Paul Johnston; Michael S. Lynch, University of Georgia; Vida Maralani, Yale University; William Mason; Georges Monette; A. M. Parkhurst, University of Nebraska-Lincoln; Doug Rivers; Paul D. Sampson, University of Washington; Corey S. Sparks, The University of Texas at San Antonio; Robert Stine; and Sanford Weisberg. I am also in debt to Paul Johnson's students at the University of Kansas, to William Mason's students at UCLA, to Georges Monette's students at York University, to participants at the Inter-University Consortium for Political and Social Research Summer Program in Robert Andersen's advanced regression course, and to my students at McMaster University, all of whom were exposed to various versions of the second edition of this text prior to publication and who improved the book through their criticism, suggestions, and—occasionally—informative incomprehension.

Edward Ng capably assisted in the preparation of some of the figures that appear in the book.

C. Deborah Laughton, Lisa Cuevas, Sean Connelly, and—most recently—Vicki Knight, my editors at Sage Publications, were patient and supportive throughout the several years that I worked on the various editions of the book.

I have been very lucky to have colleagues and collaborators who have been a constant source of ideas and inspiration—in particular, Michael Friendly and Georges Monette at York University in Toronto and Sanford Weisberg at the University of Minnesota. I am sure that they will recognize their influence on this book. I owe a special debt to Georges Monette for his contributions, both direct and indirect, to the new chapters on mixed-effects models in this edition. Georges generously shared his materials on mixed-effects models with me, and I have benefited from his insights on the subject (and others) over a period of many years.

Finally, a number of readers have contributed corrections to earlier editions of the text, and I thank them individually in the posted errata to these editions. Paul Laumans deserves particular mention for his assiduous pursuit of typographical errors. No doubt I'll have occasion in due course to thank readers for corrections to the current edition.

If, after all this help and the opportunity to prepare a new edition of the book, deficiencies remain, then I alone am at fault.

John Fox
Toronto, Canada
August 2014

About the Author

John Fox is Professor of Sociology at McMaster University in Hamilton, Ontario, Canada. Professor Fox earned a PhD in Sociology from the University of Michigan in 1972, and prior to arriving at McMaster, he taught at the University of Alberta and at York University in Toronto, where he was cross-appointed in the Sociology and Mathematics and Statistics departments and directed the university's Statistical Consulting Service. He has delivered numerous lectures and workshops on statistical topics in North and South America, Europe, and Asia, at such places as the Summer Program of the Inter-University Consortium for Political and Social Research, the Oxford University Spring School in Quantitative Methods for Social Research, and the annual meetings of the American Sociological Association. Much of his recent work has been on formulating methods for visualizing complex statistical models and on developing software in the R statistical computing environment. He is the author and coauthor of many articles, in such journals as *Sociological Methodology*, *Sociological Methods and Research*, *The Journal of the American Statistical Association*, *The Journal of Statistical Software*, *The Journal of Computational and Graphical Statistics*, *Statistical Science*, *Social Psychology Quarterly*, *The Canadian Review of Sociology and Anthropology*, and *The Canadian Journal of Sociology*. He has written a number of other books, including *Regression Diagnostics* (1991), *Nonparametric Simple Regression* (2000), *Multiple and Generalized Nonparametric Regression* (2000), *A Mathematical Primer for Social Statistics* (2008), and, with Sanford Weisberg, *An R Companion to Applied Regression* (2nd ed., 2010). Professor Fox also edits the Sage Quantitative Applications in the Social Sciences ("QASS") monograph series.